A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

1

LA-UR--83-3392

DE84 003836

CONF 831104--5

TITLE FINE TUNING OF A MEASUREMENT CONTROL PROGRAM AT THE LOS ALAMOS NATIONAL LABORATORY

AUTHOR(S) A. S. Goldman and R. S. Marshall

SUBMITTED TO ANS/INMM Conference on Safeguards Technology: The Process-Safeguards Interface, Hilton Head Island, SC, November 28-December 2, 1983

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

FINE TUNING OF A MEASUREMENT CONTROL PROGRAM
AT THE LOS ALAMOS NATIONAL LABORATORY*

A. S. GOLDMAN and R. S. MARSHALL
Los Alamos National Laboratory
Los Alamos, New Mexico 87545, USA
(505) 667-2402

ABSTRACT

This paper suggests a revised measurement control program (MCP) for balances at the Los Alamos National Laboratory plutonium facility. The revised MCP is based on an analysis of data taken from June 1981 through August 1983. The most important finding in our study is that significant measurement bias occurs in nearly every balance. An important cause of this bias has been traced to truncation errors, and a detailed discussion of the effects of truncation errors is presented. We also discuss other sources of bias and their resolution, and finally, we suggest methods for determining accuracy, precision, and randomness of measurements of weights and the response to failures of statistical tests.

INTRODUCTION

Each facility under the auspices of the Department of Energy (DOE) is required to implement and maintain a measurement control program (MCP) for all instrumentation used to measure special nuclear material (SNM). Such an MCP for balances has been operating since 1978 at the Los Alamos Plutonium Processing Facility. This paper reviews, assesses, and recommends improvements on this MCP. The study is based on 2 years of accumulated MCP data on 25 balances.

CURRENT BALANCE MCP

The Los Alamos MCP for balances conforms to specifications of DOE Order 5630.2, Section 36, which specifies that, where practicable, "All scales and balances shall be maintained in good working condition, and calibrated pursuant to an established control program." The existing Los Alamos MCP for balances[1] is based on measurements using weights traceable to NBS reference standards. Instruments cannot be used for accountability measurements unless they have passed accuracy checks within 24 hours or precision checks within 7 days. The accuracy and precision data are evaluated upon entry in a central computer and presented monthly in the form of control plots and data summaries.

The commercial electronic balances are typically 5-kg capacity and 0.1-g-readout. Calibration is done with known standards whose weights are ~1 kg and ~4 kg. These same standards are used in the 5 days/wk accuracy tests and the 1 day/wk precision tests. A description of these tests follows. (See Ref. 1 for more complete details.)

Accuracy tests are made on each balance for each weight by computing

$$Z_a = \frac{W_a - W}{0.15} ,$$

where W_a is the measured value for a standard weight (~1 kg or ~4 kg), W is the "known" value of that weight, and 0.15 g is the historical standard deviation of W_a that is applied to all balances. It is assumed that Z_a is distributed normally with zero mean and unit variance. Let the computed value of Z_a be denoted as z_a; then if $|z_a| \geq 1.96$, the accuracy test is repeated. For any test, if $|z_a| \geq 2.58$, an action message is given and the balance is repaired and/or recalibrated.

Precision tests are based on the Chi-square/degrees of freedom statistic

$$U_p = \frac{s_p^2}{0.08^2} ,$$

where s_p^2 is the sample variance of five measurements of the weight, and 0.08 g is a historical standard deviation of repeated measurements that

*Work supported by the US Department of Energy (DOE), Office of Safeguards and Security as part of an effort to improve the overall quality of MCPs at DOE facilities.

is applied to all balances. The day-to-day distribution of W_a is assumed to have a standard deviation of 0.15 g, but the distribution of measured values of a standard weight taken within a day has a standard deviation of 0.08 g. Assume that $4U_p$ has a Chi-square distribution with 4 degrees of freedom. Let u_p denote an observed value of U_p. The balance is retested if $u_p \geq 2.37$. An action message is issued if $u_p \geq 3.32$, and the balance is repaired and/or recalibrated.

This MCP has served well and has identified faulty balances since startup in 1978; however, it has some shortcomings. In the following section, these shortcomings will be explored and potential MCP improvements will be discussed.

REVISED BALANCE MCP

Our revised balance MCP includes quality control measures that would address five major problem areas.

(1) Bias. What is the bias for each measuring device and is it significantly different from zero? For a particular balance and a standard weight, bias is defined as the difference between the mean of the observed weighing and the true value of the weight.

(2) Accuracy. Is a particular weight measurement sufficiently accurate to indicate the balance is operating correctly? Accuracy is defined as the difference between an observed measure of weight and the actual weight.

(3) Precision. Do repeated weighings indicate satisfactory precision? Precision is a measure of the short-term reproducibility of a balance.

(4) Random Pattern. Do weight measurements have a random pattern during a specified time frame? In general, day-to-day weight measurements should behave in a random fashion and show neither oscillations nor trends.

(5) Failure Frequency. Is a balance failing a precision, accuracy, or random pattern test too frequently? The reason could reflect a transcription error on the part of an operator or it could indicate that the balance needs repair or recalibration.

The following discussion describes the recommended MCP that will answer the above questions. Our preliminary findings are based on a study of data obtained from 25 balances from June 1981 through August 1983. Approximately 60 balances were subjected to the current MCP; however, to improve the credibility of our results, we examined only those balances that contributed at least 90 data points. Recalibration dates were determined from log books made available by the Nuclear Material Operations Group at the plutonium facility. Information computed from the current MCP includes the number of days between recalibrations, and means and standard deviations of differences between measured weights of standards and their accepted values.

BIASES IN BALANCES

The bias for each balance is estimated by averaging the difference between measured and standard weights over a 2-year period. The estimated bias for each balance is shown in Fig. 1. The Student's "t" test was applied for each balance at the 0.05 probability level. Balances B-19 and B-20 are considered "unbiased" (zero bias) when weighing 4-kg weights, and B-07 and B-09 appear to be "unbiased" for 1-kg weights. B-09, B-16, and B-33 are borderline cases for 4-kg weights. All other balances have significant bias. The average bias for all balances is -0.02583 g for 4-kg weights and -0.0152 g for 1-kg weights.

Figure 1 shows that balance bias at the 4-kg level varies from -0.135 g (B-23) to +0.066 g (B-36). At the 1-kg level, the bias ranges from -0.106 g (B-34) to +0.093 g (B-01). Most of these biases are negative. This phenomenon has been attributed to a truncation

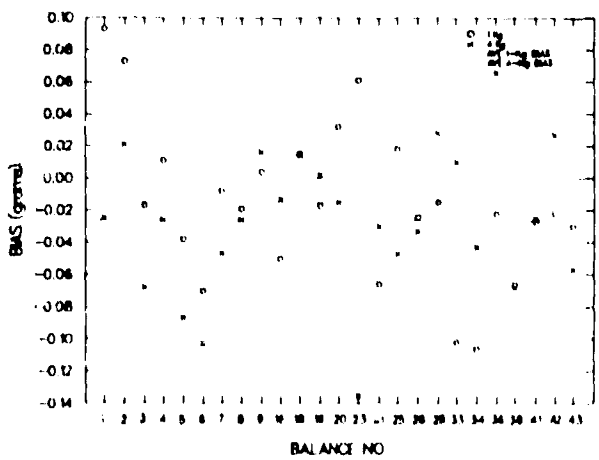

Fig. 1. Biases for 1-kg and 4-kg weights in 25 different balances.

mechanism characteristic of the balances that is explained further in the next section.

TRUNCATION EFFECT ON BIAS

Often the causes of bias are known, but estimates of contributions of these causes are unknown. In this study we know that truncation (dropping the last digit from the displayed readings) is a cause of bias, and we can estimate that on the average its contribution is -0.05 g. This section discusses ramifications of the truncation problem.

Most of the balances studied have tenth gram readout. The weight displayed is a truncated value, whereas the standard weight is known to the nearest hundredth of a gram. Differences between readings and the standard weights reflect this truncation effect. An example of the truncation effect is as follows.

Consider a balance that has a standard with weight 995.89 g. Any measured value, x, that would result in $995.80 \leq x < 995.90$ would have a readout of $x = 995.8$, and the bias would be equal to -0.09 no matter what digit occurred in the hundredth position. Thus, it is intuitive that such truncation will give negative bias. It can be shown theoretically that for balances in this study the average bias generated by truncation is $\sim$-0.05 g.

The truncation-introduced bias of $\sim$-0.05 g also can be shown empirically. Consider data from B-08 and B-09, recorded during August and September 1983, to the nearest hundredth gram. The comparison with truncated readings is given in Table I.

The results in Table I behave as expected in all four cases. Correction for truncation reduces the average bias by $\sim$0.05 g for all weights. This correction is listed as Δ in Table I and represents the difference between observed bias, relatively free of truncation effects, and bias including truncation effects.

CURRENT MCP AND BIAS

When the current MCP was established in 1978, all of the balances appeared to be performing properly; that is, there were no "bad" balances. On the other hand, it was obvious that the measurement standard deviation σ was different for the two different standards. In addition, standard deviations were different among the balances. This suggests using different standard deviations for tests on each weight and each balance. However, for convenience in implementing statistical testing procedures, the decision was made to use only one value for the standard deviation for testing purposes. The value $\sigma = 0.15$ was selected.

The truncation bias was not considered at that time. This oversight has little adverse effect on decisions made from accuracy tests because the inflated value $\sigma = 0.15$ is used.

The following three cases illustrate differences in three approaches to the accuracy test. B-23 tested with a standard weight of 4008.17 g has an observed bias of -0.135 and $\hat{\sigma} = 0.11$. Here σ is estimated by $\hat{\sigma}$, the average standard deviation over all recalibration. Suppose the measured value for the standard is 4007.8.

Case 1: Current MCP

The accuracy test would compute

$$z_a = \frac{4007.8 - 4008.17}{0.15} = -2.47 \quad ,$$

and a second test would be necessary.

Case 2: Current MCP with "correct" σ

If $\sigma = 0.11$ were used

$$z_a = \frac{4007.8 - (4008.17)}{0.11} = -3.37 \quad ,$$

leading to a recalibration.

Case 3: Suggested MCP.

Using a bias equal to -0.135 and $\sigma = 0.11$ gives

$$z_a = \frac{[4007.8 - (4008.17 - 0.135)]}{0.11} = -2.14 \quad ,$$

and a second test would be required.

The MCP test decision was the same for Case 1 (current MCP) and Case 3 (revised MCP) because of the inclusion of truncation bias in the inflated $\sigma = 0.15$.

ACCURACY

In the revised MCP it is recommended that σ estimates for individual balances (and for different standards) be used for testing. If this is not practical, we suggest 0.14 be substituted for 0.15 in the 4-kg tests and 0.09 for

TABLE I

COMPARATIVE ESTIMATES OF BIAS FOR B-08 and B-09
FOR AUGUST AND SEPTEMBER 1983

B-08

	1 kg (std = 999.55 g)			4 kg (std = 4001.57 g)		
	Bias		Δ	Bias		Δ
	(Measured)	(Truncated)	Difference	(Measured)	(Truncated)	Difference
n	32	32		32	32	
Mean	0.155	0.116	0.039	0.043	-0.004	0.047
Std. Dev.	0.052	0.060		0.089	0.097	

B-09

	1 kg (std = 996.66 g)			4 kg (std = 3995.83 g)		
	Bias		Δ	Bias		Δ
	(Measured)	(Truncated)	Difference	(Measured)	(Truncated)	Difference
n	32	32		32	32	
Mean	0.102	0.053	0.049	0.063	0.026	0.037
Std. Dev.	0.034	0.034		0.103	0.107	

1-kg tests. These values represent the largest standard deviations for each weight category as shown in Fig. 2, which is a plot of bias and standard deviation pairs for each balance.

Truncation bias can be determined and subtracted from observed standard weight values before testing for accuracy, or truncation bias can be reduced to insignificant levels by replacing balances with 0.01-g readout balances.

PRECISION

The precision test in the current MCP suffers from choosing $\sigma_p = 0.08$, a value that is too large. The standard deviation of five successive measurements, taken over a 30-min period, is so small that often the repeated measurements are identical. Figure 3 gives a control plot for B-05 in April 1983. Precision tests were run four times during the month on both 1-kg and 4-kg standard weights for a total of eight tests. Only one time out of eight are the repeated measurements nonidentical.

If standard deviations for five accuracy measurements taken on separate days of the week are calculated, $\sigma_p = 0.10$ for 4-kg tests is appropriate. We suggest that the precision tests be made on the previous five accuracy measurements rather than repeated measurements taken on a single day. This procedure would lead to a savings of operator time because the same data can be used for both accuracy and precision testing. Furthermore, because the test is made daily, this procedure reduces the risk that the balance will become erratic without detection during the week's interval between precision tests.

A sample of 566 precision tests during the period from July through September 1982 yielded only two failures. On both occasions the operator redid the test with successful outcomes. These results suggest that the current precision tests are unnecessary.

RANDOM PATTERNS

The current MCP has no test for randomness on day-to-day measurements of weights. We propose to correct this situation with a mean square successive difference test (see Ref. 2, p. 221) to be carried out on a daily basis. The sequence consists of dropping the first point and adding the last over 20 days of measurement data. Computational procedures and an example follow.

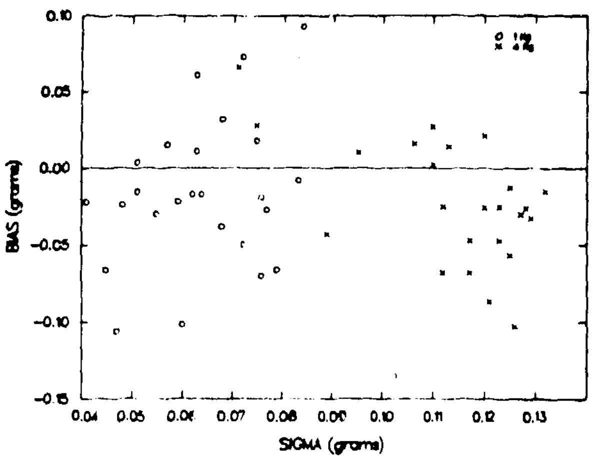

Fig. 2. Bias vs standard deviation of weight for 1-kg and 4-kg weights in 25 different balances.

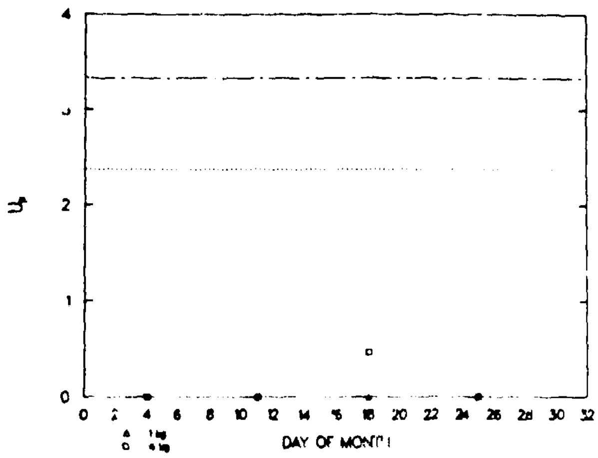

Fig. 3. Precision control plot for B-05, April 1983.

The mean square successive difference test tests the hypothesis that successive measurements over time are random in nature vs the alternative hypothesis that data are "not random" to the extent that successive differences are too small or too large. Large positive values of the statistic Z given below indicate oscillations, and large negative values indicate that each observation tends to resemble the previous observation (positive serial correlation).

The following computations are made.

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 / n}{(n-1)} ,$$

$$d^2 = \frac{\sum_{i=1}^{n-1} (x_{i+1} - x_i)^2}{(n-1)} , \quad \text{and}$$

$$Z = \frac{\frac{d^2/2}{s^2} - 1}{\sqrt{(n-2)/(n^2-1)}} ,$$

where the distribution of Z is approximated by the standard normal distribution. According to Brownlee,[2] this approximation works well for $n > 10$. If the false-alarm probability is 0.01, the hypothesis of randomness is accepted if $|z| < 2.58$; otherwise it is rejected.

A set of 20 contrived data points is given in Table II to illustrate a situation in which the mean square successive difference test can be applied. Inspection of the accuracy control plots reveals an abrupt shift from negative to positive differences between measured and standard weights. However, the data passed the current MCP tests.

The mean square successive difference test applied to the data in Table II gives

$n = 20$
$s^2 = 0.0265$
$d^2 = 0.0099$
$z = -3.82$.

The value of z is negative indicating a positive serial correlation in the data and a recalibration should be required.

RESPONSE TO FAILURES OF STATISTICAL TESTS

Recalibrations and/or repairs are made in response to repeated failures of the accuracy and precision tests. The recommendation for the revised MCP is to consider a balance as defective if the average number of days between recalibrations is <20. For convenience, the final

TABLE II

A SET OF CONTRIVED DATA POINTS

Observation Number	$X_i = W_i - W_a$	$X_{i+1} - X_i$	Observation Number	$X_i = W_i - W_a$	$X_{i+1} - X_i$
1	−0.15		11	0.19	0.22
2	−0.15	0	12	−0.05	−0.24
3	−0.20	−0.05	13	0.21	0.26
4	−0.17	0.03	14	0.22	0.01
5	−0.15	0.02	15	0.18	−0.04
6	−0.18	−0.03	16	0.17	−0.01
7	−0.14	0.04	17	0.17	0.00
8	−0.14	0.00	18	0.15	−0.02
9	−0.10	0.04	19	0.15	0.00
10	−0.03	0.07	20	0.14	−0.01

day of the study, August 31, was chosen as a calibration date for all balances. B-38, the best performer, has passed all tests for 438 successive trials. A summary of recalibration data is given in Table III. The current 4-kg accuracy test had been used to determine when recalibrations were to be made. The number of recalibrations for each balance does not seem to be unduly large except for B-05. If the 20-day rule were put into effect, B-05 and B-24 would be under very close scrutiny. The last recorded recalibration for B-05 was on August 15, 1983. Before April 7, 23 recalibrations were required for an average of nearly 13 days between recalibrations.

Although the 20-day rule is not part of the current MCP, B-06 averaged 19 days between recalibrations and was replaced in April 1982.

A frequency of failure test program needs to recognize that the calibration of a balance requires special skills. This was taken into consideration in Table III. We considered recalibration on successive days as an operator error and excluded such data. For example, B-01 had three recalibrations within 2 days and was counted as only one recalibration.

Some balances are difficult to adjust and appear worse than they really are. For instance, B-43 had six recalibrations in a short period of time but then went 64 days before its next recalibration; B-43 is probably a good balance.

In spite of the practical difficulties experienced in reviewing large numbers of control plots, it is recommended that control plots be continued as an operational tool in observing accuracy and precision test patterns. Control plots for accuracy tests on B-05 during the month of April 1983 are shown in Fig. 4. There are five points outside the boundary $|z_a| > 2.58$, indicating that the accuracy test failed

TABLE III

SUMMARY OF RECALIBRATIONS
JUNE 1981 THROUGH AUGUST 1983

Balance	Number of Recalibrations	Total Operating Days	Average No. of Days Between Recalibrations
B-01	18	381	21.17
B-02	9	326	36.22
B-03	4	399	99.75
B-04	12	320	26.67
B-05	24	312	13.00
B-06	8	142	17.75
		(Scale removed)	
B-07	8	308	38.50
B-08	6	265	44.17
B-09	7	189	27.00
B-16	5	384	76.80
B-18	3	327	109.00
B-19	0	411[a]	
B-20	11	278	25.27
B-21	21	201	9.57
		(Scale Replaced)	
B-22	4	396	99.00
B-24	15	288	19.20
B-25	8	215	26.88
B-26	4	348	87.00
B-28	0	406[a]	
B-33	7	400	57.14
B-34	3	406	135.33
B-36	3	368	122.67
B-38	0	436[a]	
B-41	10	373	37.30
B-42	14	330	23.57
B-43	13	346	26.62

[a]Days balance operated up to August 31, 1983, and was never recalibrated.

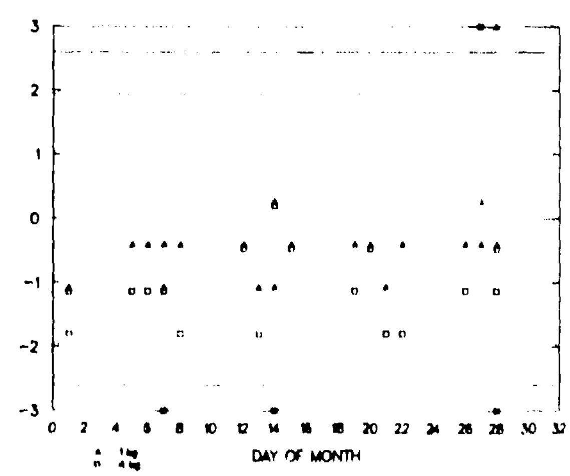

Fig. 4. Accuracy control plot for B-05, April 1983.

an incredible five times. An investigation revealed that a new operator on duty had difficulty in adjusting the balance.

SUMMARY

The present and suggested MCPs are summarized in Table IV. Finally, it is recommended that only balances that give readings to hundredths of a gram be used.

ACKNOWLEDGME ~

Computations for this report were made in a computer program written by M. L. Bonner of the Safeguards Systems Group (Q-4). Helpful discussions with J. T. Markin, W. J. Conover, and R. G. Gutmacher were instrumental in making improvements on this report. Unpublished reports by C. C. Thomas and J. Sapir provided helpful background material.

REFERENCE

1. W. Severe and C. Thomas, "Measurement Control Program for In-Line NDA Instruments," Nucl. Mater. Manage. VIII (Proceedings Issue) 620-633 (1979).

2. K. A. Brownlee, Statistical Theory and Methodology in Science and Engineering, 2nd Ed. (John Wiley and Sons, Inc., New York, 1965).

TABLE IV

A SUMMARY OF THE CURRENT AND SUGGESTED MCPs

Bias	Accuracy Test	Precision Test	Randomness Test	Failure Frequency
		Current MCP		
No computation of bias made	For each measurement: 1. Compute: $Z_a = \frac{W_a - W}{0.15}$ 2. Pass if $\lvert Z_a \rvert \leq 1.96$ Repeat if $1.96 < \lvert Z_a \rvert$ Recalibrate or repair if $\lvert Z_a \rvert > 2.58$ W_a: measured weight W: known standard (~1 kg or ~4 kg)	Each week for each balance 1. Compute: $u_p = \frac{s^2}{0.08}$ s^2: variance of 5 repeated measurements 2. Pass if $u_p < 2.37$ Repeat if $2.37 \leq u_p$ Recalibrate or repair if $u_p > 3.32$	No test	Subjectively replace balance if too many failures occur
		Suggested MCP		
Compute bias b_j for each balance and each standard $b_j = \frac{\sum_{i=1}^{n_j} (W_{ai} - W_j)}{n_j}$ where W_{ai} is the i^{th} measurement since the most recent calibration, W_j is the known weight, and n_j is the number of observations, respectively, for the j^{th} balance	For each measurement: 1. Compute $\hat{\sigma}_j$, the average standard deviation over previous recalibrations 2. Compute bias b_j 3. Compute $Z_a = \frac{W_a - W - b_j}{\hat{\sigma}_j}$ See step 2 in current MCP	Use 5 daily measurements instead of 5 repeated, and use 0.10 instead of 0.08. If test fails, recalibrate or repair	Use mean square successive difference test described in text	Investigate if average number of days between recalibration exceeds 20